SUPPLEMENTARY INFORMATION

The transcriptomic context of DRD1 is associated with prefrontal activity and behavior during working memory

INVENTORY OF SUPPLEMENTARY INFORMATION [SI]

1. SI MATERIALS AND METHODS

WGCNA PREPROCESSING

We included probes in the categories human constitutive exonic (i.e., probes tagging exons common to multiple isoforms), human alternative exonic (i.e., probes tagging isoform-specific exons), and human mRNA (i.e., tagging untranslated regions of gene-coding mRNA). When genes were represented by multiple probes highly correlated with each other (Bonferroni-corrected $\alpha < .01$, i.e. Pearson's $r = .36$), we selected the ones with the greatest variance, following previous work (1). We factored out the effects of confounding variables from each gene's expression using a stepwise linear model, selected via the Akaike Information Criterion and applied a rank-based inverse normal transformation to reduce the impact of outliers and deviations from normality (2). Thus, we obtained a 199 × 23,636 matrix of expression residuals reflecting for each subject transcription levels of all genes relative to the entire sample considered. These variables were used for the identification of an unsigned co-expression network (3). WGCNA detected non-overlapping sets of co-expressed probes, called modules, based on the analysis of gene-gene Pearson's correlation indices. The correlation matrix was raised to an exponent β selected to guarantee scale invariance (3, 4). Scale invariance is reckoned as an important property of biological networks. A scale invariant network includes key genes, called hubs, with many connections, and more peripheral genes with selective connections. Importantly, a scale invariant network can be examined at different granularities, i.e., one can focus on larger or smaller groups of genes (called modules) based on the parameters used to identify the network. For the network at hand, $\beta = 4$ guaranteed scale invariance ($R2 = .86$; see (3)); minimum gene set size was 40 genes and minimum height for merging gene sets was .05.

PARCELING PROCEDURE

DRD1 clustered within a module including 2452 probes, a large number difficult to study as a single unit. This module could not be parceled even when lowering cut height down to .01 (note that WGCNA default is .15). Since the WGCNA network has, by definition, a nested structure with highly connected hub genes and a modular organization, it should be possible to parcel a module into smaller gene sets like the transcriptome is parceled into modules by WGCNA. Larger gene sets will include also genes with weak relationships, while smaller sets will eventually yield a loss of information. Parceling a gene co-expression module requires the definition of a target variable for optimizing the outcome. Therefore, we focused on information content associated with the module, as indexed by entropy. Entropy H is intrinsically related to the information content provided by a certain partition, as information I is, by definition:

$$I = H\_max - H$$

where H_max is the maximum entropy of a system with the same number of elements. For a weighted complex network:

$$H\_max = W \log2\ W$$

with W being the total amount of the network weights. Given the above definition, entropy and information are inversely proportional up to an additive constant which is the maximum entropy of the observed system. Thus, we iteratively parceled the module until we identified the *DRD1* gene set characterized by the maximum information content indexed in terms of entropy of the nodes at multiple parceling steps. At each parceling step, we used betweenness-based thresholding - i.e., a centrality measure of a given gene based on how many of the connections in each module are mediated by that gene (5) - to cluster genes into sets, and then we iterated the procedure on the gene set that included *DRD1*. The parceling procedure allowed us to identify the *DRD1* gene set, within the original WGCNA module, with the maximum average betweenness.

The optimization aims to obtain a trade-off between two opposite trends: if an optimal partition exists, it should be outlined by the presence of a local maximum which identifies a stronger and more stable network of genes around *DRD1*; however, as the cardinality of selected genes decreases when iteratively parceling the *DRD1* gene set, the informative content should decrease, as a smaller number of genes belong to the set. Thus, the iterative parceling of the *DRD1* set is expected on the one hand to outline a cluster of genes which carry the maximum information, and on the other hand to show that further partitions cannot yield any substantial improvement.

STATISTICAL ANALYSES ON THE DRD1 GENE SET

To understand how *DRD1* expression was related with gene set expression, we computed the Pearson's correlation of *DRD1* expression with the first principal component of gene set expression (gene set eigengene, GSE). The GSE was not significantly associated with any of the confounders considered in the preprocessing (age, sex, ethnicity, RIN, pH, post-mortem interval; all $p > .05$), suggesting successful noise removal. Additionally, we tested whether the expressions of *DRD1* and *DRD2* were correlated in Braincloud using Spearman's ρ.

In BrainEAC, we selected high quality observations with RIN ≥ 6 (N = 26; note that all Braincloud and CMC samples had RIN ≥ 7). The CMC sample included 179 healthy subjects of Caucasian or African American ancestry, RNA integrity number (RIN) ≥ 7 (see Fromer et al. (6) for details on sample collection and characteristics). We filtered out transcripts with median Reads Per Kilobase per Million mapped reads (RPKMs) ≤ 0.1 and 107 of the 126 genes in the *DRD1* gene set survived. We used the R package RUVcorr to remove noise from the replication gene expression datasets. RUV (removal of unwanted variation) derives estimates of systematic noise directly from the data

and is ideally suited for gene co-expression analyses. In BrainEAC and in CMC, we used RUV to extract five latent confounding variables and marginalized gene expression values for them.

We investigated whether gene expression covariation in both replication datasets had the same direction observed in Braincloud. To this aim, we plotted gene by gene the factor loadings obtained via Principal Component Analysis in the discovery and replication datasets. When a gene had a positive or negative loading in both datasets was defined as 'concordant in sign' while in the other cases 'discordant in sign'. We used two-sided binomial exact tests to assess whether genes 'concordant in sign' outnumbered genes 'discordant in sign' (p-value < .05).

SNP ASSOCIATION STUDY

We used SNPs & Variation Suite (SVS, GoldenHelix, Bozeman, Montana) to select SNPs associated with the gene set eigengene (GSE). We restricted our analyses to SNPs whose genotypes were available both in Braincloud and in the samples recruited for the behavioral and fMRI studies (546.308 SNPs in common out of 654.333 Braincloud SNPs). SNPs significantly deviating from Hardy-Weinberg equilibrium ($\alpha = .003$; (7)) or with a Minor Allele Frequency (MAF) $\leq .1$ were excluded (3) because the sample size was too limited to investigate rare variants. Furthermore, we reasoned that a SNP associated with co-expression, but not associated with the expression of any single gene, would possibly be a spurious result. Thus, we aimed to increase the biological plausibility of the findings by filtering out SNPs that were not statistically associated with any of the genes in the set at nominal p-value ($\alpha = .05$).

To associate SNPs with the GSE, we used ANOVAs with jackknife resampling in a leave one-out framework. In other words, we used 199 samples including 198 subjects by leaving out one

subject per run; for each of the 199 samples, we computed one ANOVA per SNP. Then, we repeated the procedure for each SNP and each sample and obtained 199 p-values of the GSE association with each SNP. Finally, we computed the median p-value per SNP and ranked SNPs by this p-value. The present procedure is more robust compared to our previous report (3) because we performed additional resampling to reduce the influence of extreme expression values. Moreover, we asked whether these SNPs were associated with gene expression in their own locus and with *DRD1* expression. To this aim, we computed ANOVAs using the selected SNP genotypes as independent factors and the expression of the closest gene of the set as the dependent variable. We also assessed the biological significance of the SNP set identified by interrogating the software Haploreg (http://archive.broadinstitute.org/mammals/haploreg). Haploreg is a dataset of genetic regulatory elements, including cis- and trans-eQTL hits across multiple tissues according to previous genomic studies (8-10).

After SNP selection, we computed the Polygenic Co-expression Index (PCI). This step is needed to translate the knowledge we gained via data mining on post-mortem samples into brain activity prediction in living subjects. In other words, we used genetic variants to compute a proxy of individual *DRD1* gene set co-expression which we called *DRD1*-PCI. For each SNP, we used the major allele homozygote (MH) as the reference group. We computed a distance measure for the heterozygous and the minor homozygous samples relative to the MH group (7). After having defined the effect of each genotypic population within individual SNPs, we defined the PCI of each participant as the arithmetic mean of genotypic effects. We used a slightly modified computation of the genetic index compared to previous works, i.e., we used the A' index instead of the d' because A' can be computed without assuming a normal distribution of the variables (11, 12).

We tested whether the PCI was associated with any of the confounders considered in the preprocessing (age, sex, ethnicity, RIN, pH, post-mortem interval) by means of Pearson's

correlations and independent samples t-tests. To assess the predictive power of the PCI, we used cross-validation. We were aware that 10-fold cross-validation is the popular framework for model selection (13, 14). Nonetheless, leave-one-out cross-validation tends to yield unbiased estimations (at a cost of high variance) and with the current sample size it could be preferred with a limited risk of overfitting. In particular, (i) at each cycle of the cross-validation, we computed the leave-one-out GSE in 198 out of 199 subjects; (ii) we ranked SNPs based on the p-value of their association with the GSE; (iii) we retained the top 13 SNPs, i.e., the same number of SNPs included in our PCI. We chose this method because imposing a p-value threshold would vary, for each iteration, the amount of information (i.e., the number of SNPs) and the range of the polygenic score (polygenic scores with more SNPs tend to have smaller ranges), ultimately adding noise to the procedure; (iv) we computed SNP weights and used them to compute the *DRD1*-PCI in the left-out subject; (v) we computed Pearson's correlation of the cross-validated *DRD1*-PCI in the test set (i.e., the ensemble of left-out subjects) with the GSE as well as with *DRD1* expression (Supplementary Figure 2). We used a meta-analytic approach to compare the effect size of the GSE prediction in the three datasets employed. We computed a fixed effect meta-analysis over Braincloud, BrainEAC, and CMC using the *metafor* R package. For the discovery set, Braincloud, we used leave-one out cross-validation (LOOCV) to estimate unbiased effect size. Cochran's Q-test for heterogeneity served to assess the combinability of studies. Finally, we used the cross-validated data also to assess the effect of ethnicity on the relationship between the *DRD1*-PCI and the GSE by means of an ANCOVA.

IMAGING AND BEHAVIORAL STUDY

Inclusion criteria and socio-demographic assessment

All participants were assessed for IQ (WAIS-R) (13), handedness (Edinburgh Inventory) (14), and socio-economic status (Hollingshead Four Factor Index of Social Status) (15). All individuals were

evaluated with the Structured Clinical Interview (16) for the Diagnostic and Statistical Manual of Mental Disorders to rule out current or past psychiatric disorder. Other exclusion criteria were first-degree familiarity for major psychiatric disorders, history of drug or alcohol abuse, active drug use in the past year, brain alterations or illness as evaluated by a board- certified neuroradiologist (TP), head trauma with loss of consciousness, and any relevant medical condition. Only participants whose behavioral accuracy was significantly above chance were included in the study, as detailed by Pergola et al. (3). In particular, given the number of trials in each session, we derived a 95% confidence interval for chance level performance (asymptotic threshold: accuracy = 38% for the 3-back fMRI study, which we applied to all WM conditions). Accuracy ≥ 38% could be statistically discriminated from chance level and was an inclusion criterion for the study. Additionally, only individuals whose scans were not affected by scanning artifacts or excessive movement were included in the study.

Working memory task

The term "N-Back" refers to the number of items - back in the sequence of stimuli - participants were required to maintain to perform the task. The stimuli consisted of numbers shown in random sequence and displayed at the points of a diamond-shaped box. There was a non-memory control condition (0-Back) that simply required subjects to identify the stimulus currently seen. In the WM condition, the task required the recollection of a number seen one (1-Back), two (2-Back) or three stimuli (3-Back) before, while continuing to encode additionally incoming stimuli. The stimuli were arranged in 30 s blocks including 2 s of instructions and 14 task trials lasting 2 s each (stimulus presentation: 500 ms, inter-stimulus interval of 1500 ms). During fMRI, we used a block design, consisting of four blocks of the control condition alternating with four blocks of each WM condition. Participants did not exit the scanner during the breaks. Stimuli were shown via a back-

projection system and behavioral responses were recorded through a fiber-optic four-button MRI-compatible pad within the scanner. Participants responded with their right hand with a four-button MRI-compatible pad within the scanner.

In the behavioral studies, stimuli were presented on a computer monitor and participants responded with their right hand on the numeric keypad of a standard computer keyboard.

fMRI data acquisition and analysis

Blood oxygen level-dependent (BOLD) signal was recorded by a GE Signa 3T scanner (General Electric, Milwaukee, WI), using a gradient-echo planar imaging sequence (repetition time, 2000 ms; echo time, 28 ms; 20 interleaved axial slices; thickness, 4 mm; gap, 1 mm; voxel size, 3.75 × 3.75 × 5 mm; flip angle, 90°; field of view, 24 cm; matrix, 64 × 64). The first four scans were discarded to allow for equilibration effect. Imaging data quality was ascertained as detailed elsewhere (3).

We realigned the images to correct for motion artifacts using the Realign and Unwarp function (SPM12). Movement parameters were extracted to exclude data affected by excessive head motion (2.5 mm of translation or 2.5° of rotation). Realigned images were resliced to a 3.75 mm isotropic voxel size, spatially normalized into a standard space (Montreal Neurological Institute) by using a 12-parameter affine model, and smoothed with a 10 mm full-width at half-maximum isotropic kernel. Thus, 24 regressors of movement were added in the statistical model using the Friston24 model (17). Furthermore, a binary regressor of no-interest, indicating volumes with excessive movement (scan-to-scan motion > 1mm) was included in the model. After spatial pre-processing, we estimated one boxcar model convolved with the hemodynamic response function per condition. Then, effects of condition at individual level were evaluated producing a t statistical map for the contrasts of interest (1-back versus 0-back, 2-back versus 0-back and 3-back versus 0-back). These individual contrast images were used for group-level analyses.

In the fMRI discovery sample, we tested the association of the *DRD1*-PCI with brain

activation using a general linear model (within-subject factor: load [1-back, 2-back, 3-back]; between-subject factor: gender; continuous predictors: linear and quadratic terms of the *DRD1*-PCI and of the *DRD2*-PCI; covariates: age and socio-economic status, -which was associated with the PCI-). In the fMRI replication sample, we computed an ANCOVA on the 2-Back > 0-Back contrast (between-subject factor: gender; continuous predictors: linear and quadratic terms of the *DRD1*-PCI; covariates: age and socio-economic status; whole brain FWE corrected $p<.05$; cluster extent threshold=5).

Behavioral analyses

In the behavioral discovery sample, we computed the differential accuracy between 2-Back and 1-Back (Δ_{2-1}) and between 3-Back and 2-Back (Δ_{3-2}) (18). Thus, we computed a repeated measures general linear model (within-subject factor: load [Δ_{2-1}, Δ_{3-2}]; between-subject factor: gender; continuous predictors: linear and quadratic terms of the *DRD1*-PCI; covariates: age, and socio-economic status; $\alpha=.05$, two tailed). For the behavioral replication sample, we computed a general linear model with Δ_{2-1} as the dependent variable, gender as the between-subject factor, the linear and quadratic terms of the *DRD1*-PCI and of the *DRD2*-PCI as continuous predictors, as well as age and socio-economic status as covariates ($\alpha=.05$, one tailed).

Assessment of the relationship between the DRD1-PCI and socio-demographics

After genotyping, the *DRD1*-PCI was computed for all subjects. We included age and gender as nuisance covariates in all analyses to account for demographic variability within and between samples. We assessed the associations between the *DRD1*-PCI and possible confounding variables (socio-economic status, handedness and IQ; $\alpha=.1$) in all of the samples using Pearson's correlation analyses. Variables correlated with the *DRD1*-PCI in either of the samples were included in all

statistical models in the behavioral and imaging investigations to prevent misattributing the effect of confounding variables to the *DRD1*-PCI.

MEDIATION ANALYSIS ON WM PHENOTYPES

We designed a mediation model to assess whether the Δ_{WM}-*DRD1*-PCI relationship was direct or mediated by brain activity; a separate mediation/moderation model assessed whether the *DRD2*-PCI moderated that relationship. Both models are reported in Supplementary Figure 5, which also reports the weights obtained in the discovery sample. We used the SPSS toolbox PROCESS (19) to run the mediation analyses. Brain activity was indexed by voxels in which activity was associated with both PCIs (negative for the *DRD1*-PCI, positive for the *DRD2*-PCI; null conjunction, $p<.05$). We identified a left and a right prefrontal clusters (x, y, z: 38, 53, 20; x, y, z: - 37, 49, 20) from which we extracted BOLD signal change using MarsBaR (http:\\marsbar.sourceforge.net). We averaged the signal change in these two ROIs because we did not hypothesize different mediation paths for the left and right hemisphere. The linear term of the *DRD1*-PCI was the independent variable, mean Δ_{WM} across loads was the dependent variable, PFC BOLD signal change was the mediator (first model). In the second model, both linear and quadratic *DRD2*-PCI were moderators of the relationship between the *DRD1*-PCI and brain/behavioral phenotypes. Age, gender, socio-economic status and the quadratic term of the *DRD1*-PCI were entered as nuisance covariates.

We replicated the effects found significant in the discovery set in additional mediation/moderation models computed in the fMRI replication set. Also in this dataset, we indexed brain activity using a cluster associated with both PCIs (null conjunction, $p<.05$; x, y, z: - 33, 30, 46). In both samples, data were standardized before being entered in the model. We used bootstrap estimates of the confidence intervals (5000 resamplings) to confirm the significance of the results.

2. SI RESULTS

PARCELING

The iterative procedure we employed reached a plateau of maximum information content after four runs. More in detail, we observed that the information content reached a plateau, meaning that further parceling could not yield any increment of the informative content of the selected community. Besides, as information depends on the cardinality of the community, we also normalized entropy with the number of nodes and observed that average entropy reached a sound plateau (Supplementary Figure 1).

SNP ASSOCIATION STUDY AND CO-EXPRESSION POLYGENIC INDEX COMPUTATION

A set of 3717 SNPs within 100 kbp from the genes included in the *DRD1* gene set that we identified had genotypes available in Braincloud, BrainEAC and in the GWAS dataset of our volunteers. Of these SNPs, 3079 were associated with the expression of at least one gene of the set ($p < .05$). After jackknife resampling, we obtained a set of 13 independent SNPs associated with the GSE with median resampled $p < .005$. None of these SNPs have been previously reported in association with imaging or cognitive phenotypes. Cross-validation and replication datasets supported the reliability of the prediction (Supplementary Figure 2). The estimated meta-analytical effect size was $r = -0.156$ (95% confidence intervals: -0.25 – -0.06; $p = 0.0016$). The three studies were not significantly heterogeneous (Cochran's $Q = 1.13$, p-value = 0.57). Supplementary Figure 3 shows that the effects in the cross-validated Braincloud data were not driven by one of the ethnicities, but were uniform across both.

SOCIODEMOGRAPHIC AND BEHAVIORAL DATA

Correlation analyses revealed marginally significant associations between socio-economic status and both the linear (R = .15, p = .074) and quadratic (R = -.14, p = .086) terms of the *DRD1*-PCI only in the fMRI/behavioral discovery sample. Therefore, socio-economic status was included in all tests as a nuisance variable. All other tests yielded no significant results (all p>.1). In addition to the effect of the *DRD1*-PCI, the behavioral analysis on the discovery sample revealed main effects of age ($F_{1,144}$ = 5.6; p = .018; partial η^2 = .038), and gender ($F_{1,144}$ = 4.6; p = .033; partial η^2 = .031) on Δ_{WM}. Higher age and female gender were associated with lower Δ_{WM} scores.

MEDIATION ANALYSIS

Supplementary Figure 5 illustrates the models used and the results of the discovery sample. The total effect (direct and indirect) of the *DRD1*-PCI on Δ_{WM} was significant (standardized coefficient = .21; p = .014). The relationship between *DRD1*-PCI and Δ_{WM} was largely mediated by prefrontal activity: while the direct effect was marginally significant (standardized coefficient = .16; p = .063; 95% confidence interval [-.0083; .32]; bootstrap interval [.0022; .34]), the indirect effect was significantly different from zero (standardized coefficient = .051; 95% confidence interval [.0038; .11]; bootstrap interval [.0022; .14]). However, none of these effects were significant in the fMRI replication sample (all p > .05).

When taking into account moderation effects by the *DRD2*-PCI, the indirect effect of the *DRD1*-PCI remained significant but was not significantly moderated (Supplementary Figure 5); the direct effect of the *DRD1*-PCI was significant and was moderated by the quadratic *DRD2*-PCI. Also in this case, the effects were not significant in the replication sample (all p > .05).

REFERENCES

1. **Roussos P, Katsel P, Davis KL, Siever LJ, & Haroutunian V (2012) A system-level transcriptomic analysis of schizophrenia using postmortem brain tissue samples. *Archives of general psychiatry* 69(12):1205-1213.**

2. **Beasley TM, Erickson S, & Allison DB (2009) Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior genetics* 39(5):580-595.**

3. **Pergola G, *et al.* (2017) DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. *Translational psychiatry* 7(1):e1006.**

4. **Zhang B & Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4:Article17.**

5. **Barrat A, Barthelemy M, Pastor-Satorras R, & Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci U S A* 101(11):3747-3752.**

6. **Fromer M, *et al.* (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* 19(11):1442-1453.**

7. **Pergola G, *et al.* (2016) Combined effect of genetic variants in the GluN2B coding gene (GRIN2B) on prefrontal function during working memory performance. *Psychological medicine* 46(6):1135-1150.**

8. **Westra HJ, *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 45(10):1238-1243.**

9. **Lappalainen T, *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501(7468):506-511.**

10. **Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235):648-660.**

11. **Snodgrass JG & Corwin J (1988) Pragmatics of measuring recognition memory: applications to dementia and amnesia. *J Exp Psychol Gen* 117(1):34-50.**

12. **Pollack ES & Norman VB (1964) A non-parametric analysis of recognition experiments. . *Psychonomic Science* 1.**

13. **Orsini A & Laicardi C (2000) Factor structure of the Italian version of the WAIS-R compared with the American standardization. *Perceptual and motor skills* 90(3 Pt 2):1091-1100.**

14. **Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1):97-113.**

15. **Hollingshead AB & Redlich FC (2007) Social class and mental illness: a community study. 1958. *American journal of public health* 97(10):1756-1757.**

16. **First MB, Gibbon M, Spitzer RL, & Williams JBW (1996) Guide for the structured clinical interview for DSM-IV axis I disorders-Research version., (Biometrics Research, New York).**

17. **Friston KJ, Williams S, Howard R, Frackowiak RS, & Turner R (1996) Movement-related effects in fMRI time-series. *Magn Reson Med* 35(3):346-355.**

18. **Cassidy CM, *et al.* (2016) Dynamic Connectivity between Brain Networks Supports Working Memory: Relationships to Dopamine Release and Schizophrenia. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 36(15):4377-4388.**

19. **Hayes AF (2012) PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling. *[White paper]*.**

Supplementary Table1. Genes in the *DRD1*gene set. The expression of genes in bold font is negatively correlated with the gene set eigengene. All gene-eigengene correlations are significant ($p \leq 1.6 \times 10^{-7}$).

Gene	ID Braincloud	ID BrainEAC	Chr	Strand	Start	Stop	Correlation with gene set eigengene (Pearson's r)
ARAP1	hHC017995	t3381241	chr11	-	72395373	72463414	0.66
ATF5	hHC020747	t3839103	chr19	+	50431393	50437188	0.68
ATL2	hHC023382	t2548776	chr2	-	38521117	38662637	-0.75
BCL7C	hHR022997	t3688038	chr16	-	30845382	30905613	0.75
CAPZA1	hHR026867	t2352228	chr1	+	113144301	113214213	-0.69
CCDC81	hHC017139	t3343293	chr11	+	86085803	86134148	0.64
CCNY	hHC025213	t3242425	chr10	+	35535804	35860851	0.66
CCNYL1	hHC029004	t2525182	chr2	+	208576254	208626563	-0.64
CCPG1	hHA033972	t3625326	chr15	-	55632255	55700706	-0.59
CHD4	hHC022301	t3442054	chr12	-	6679260	6716806	0.75
CHD8	hHR011006	t3528078	chr14	+	21853383	21853961	0.76
CHIC1	hHR002375	t3981609	chrX	+	72782993	72906935	-0.78
CHTF8	hHR027714	t3696454	chr16	-	69151920	69166482	0.71
CPT1B	hHC016042	t3966057	chr22	-	51007298	51021504	0.61
CRKL	hHC019229	t3937787	chr22	+	21271699	21308024	0.43
CRLS1	hHA034867	t3875242	chr20	+	5986743	6020695	-0.72
CTRL	hHC020947	t3695983	chr16	-	67962417	67965755	0.56
CUL4B	hHR004728	t4019900	chrX	-	119658474	119751514	-0.87
CUL9	hHC019676	t2907754	chr6	+	43149932	43192318	0.69

DMBX1	hHC018003	t2334847	chr1	+	46964435	46981890	0.47
DNAJB2	hHC022495	t2528426	chr2	+	220144061	220152056	0.77
DRD1	hHC010798	t2888010	chr5	-	174867047	174914186	-0.61
ENGASE	hHR015660	t3736666	chr17	+	77068718	77084674	0.55
ERAF	hHC023223		chr16	+	31446686	31447624	0.51
FAM171A1	hHR009740	t3279154	chr10	-	15212428	15416551	0.36
FAM69A	hHR010901	t2423175	chr1	-	93298463	93427025	-0.83
FBF1	hHR023754	t3771154	chr17	-	73905660	73905777	0.61
FBXO11	hHC013818	t2552153	chr2	-	48016104	48134792	-0.83
GABBR1	hHR013309	t2947889	chr6	-	29556265	29611309	0.68
GAD1	hHA035071	t2514969	chr2	+	171670391	171717661	-0.87
GAK	hHC015505	t2756673	chr4	-	843110	926149	0.73
GCN1L1	hHC010736	t3474256	chr12	-	120565014	120632632	0.71
GFER	hHC024889	t3644297	chr16	+	2034155	2037747	0.74
GLIPR1	hHC003812	t3422855	chr12	+	75874480	75895696	-0.63
GNASAS	hHC012579	t3911644	chr20	-	57393838	57425958	0.51
GUF1	hHC005703	t2725779	chr4	+	44680454	44706387	-0.81
HSD3B7	hHC019388	t3656737	chr16	+	30996532	31000455	0.66
HSF4	hHC022343	t3665215	chr16	+	67193888	67198077	0.78
IDUA	hHC014618	t2714407	chr4	+	966780	1008579	0.76
INPPL1	hHC021849	t3339423	chr11	+	71934262	71950148	0.47
IPO4	hHC017939	t3557851	chr14	-	24649263	24658104	0.47
KIF21A	hHA036028	t3450775	chr12	-	39687033	39853696	-0.88
KLHDC8B	hHC020742	t2622006	chr3	+	49207238	49213918	0.41
KLK3	hHA039893	t3839538	chr19	+	51358171	51365366	0.45
KRT8	hHR031337	t3455516	chr12	-	52914520	53343645	0.82
LRBA	hHC015424	t2789266	chr4	-	151185606	151936869	0.44
LYRM7	hHA035681	t2828115	chr5	+	130506483	130538094	-0.86

MAFK	hHR018888	t2987199	chr7	+	1550088	1580520	0.37
MAN2B1	hHC012345	t3851545	chr19	-	12757336	12777828	0.74
MAP2K3	hHC011325	t3714729	chr17	+	21187988	21218520	0.69
MAP3K3	hHC019533	t3730806	chr17	+	61699811	61773660	0.79
MGAT4A	hHC020284	t2566414	chr2	-	99233700	99347568	-0.64
MIA3	hHC031255	t2381865	chr1	+	222840225	222842596	-0.74
MKS1	hHC020431	t3764199	chr17	-	56282803	56296885	0.75
MMP15	hHC019951	t3663074	chr16	+	58056585	58145498	0.65
MNAT1	hHA035592	t3538703	chr14	+	61201460	61435480	-0.39
MOV10	hHC019728	t2352275	chr1	+	113215783	113245690	0.62
MYH9	hHR022934	t3959451	chr22	-	36677332	36784036	0.56
NDST1	hHC026687	t2835531	chr5	+	149865285	149963329	0.50
NEGR1	hHR003325	t2418078	chr1	-	71868628	72748513	-0.60
NSD1	hHC017145	t2842951	chr5	+	176545573	176724414	0.73
NUMA1	hHC019087	t3380901	chr11	-	71712261	71791719	0.78
OR1Q1	hHR018965	t3188186	chr9	+	125372029	125381771	0.47
OSBPL8	hHC005823	t3462949	chr12	-	76745119	76953589	-0.45
PCNX	hHC018743	t3542689	chr14	+	71374122	71666427	0.42
PGPEP1	hHC023256	t3824963	chr19	+	18451431	18477726	0.62
PHF20L1	hHA035883	t3116535	chr8	+	133787604	133871286	-0.77
PIK3R1	hHA033950	t2813060	chr5	+	67272645	67597647	-0.82
PLA2G15	hHC018750	t3666124	chr16	+	68279253	68294958	0.67
PLEKHA1	hHR003443	t3268274	chr10	+	124134148	124191852	-0.85
PLOD1	hHC015919	t2320581	chr1	+	11994287	12035587	0.74
PLSCR3	hHC019988	t3743701	chr17	-	7291950	7298142	0.68
POU2F2	hHC026303	t3863435	chr19	-	42585012	42636615	0.68
PPP1CB	hHA035365	t2475209	chr2	+	28974624	29025796	-0.86
PPP4R4	hHA032931	t3549605	chr14	+	94640649	94746072	-0.83

PPP4R4	hHA034126	t3549605	chr14	+	93710427	93815825	-0.82
PSIP1	hHA036176	t3199790	chr9	-	15460303	15576931	-0.88
PSIP1	hHA034232	t3199790	chr9	-	15454065	15500989	-0.73
PSIP1	hHA040058	t3199790	chr9	-	15454065	15500989	-0.83
PSME1	hHA038570	t3529609	chr14	+	24604814	24608139	0.61
PSME2	hHC032217		chr14	-	23682414	23686270	0.63
QSOX1	hHC017919	t2369950	chr1	+	180086444	180169860	0.80
RCC2	hHC031015	t2398894	chr1	-	17733276	17787445	0.68
RECQL4	hHC026358	t3158767	chr8	-	145736675	145743245	0.66
ROR2	hHC015909	t3214496	chr9	-	94325383	94712434	0.59
RPS6KA2	hHR008402	t2984655	chr6	-	166809678	167328112	0.66
RPS6KB1	hHC031266	t3729294	chr17	+	57970457	58074323	-0.58
RTN3	hHA034196	t3333942	chr11	+	63448921	63528015	-0.73
SBDSP	hHR027838		chr7	+	71744736	71749236	-0.66
SCAMP3	hHC015426	t2437307	chr1	-	155225782	155232195	0.69
SCOC	hHA040719	t2744980	chr4	+	141105311	141328721	-0.91
SCRIB	hHA039524	t3157751	chr8	-	144873106	144897533	0.60
SDCBP	hHA033272	t3099750	chr8	+	59442654	59496406	-0.83
SEC22A	hHA034208	t2639309	chr3	+	122920747	122994410	-0.71
SFI1	hHC025436	t3942998	chr22	+	31891926	32014532	0.70
SFRS16	hHC024609	t3835983	chr19	+	45542297	45574205	0.79
SIX5	hHC015139	t3865618	chr19	-	46268046	46272312	0.69
SLC25A40	hHC003156	t3060245	chr7	-	87462897	87505669	-0.88
SLC26A6	hHC018185	t2673547	chr3	-	48663166	48672953	0.65
SLC6A15	hHA035353	t3464276	chr12	-	85253267	85338161	-0.85
SLC6A8	hHR029029	t3688878	chr16	-	32797028	32896822	0.78
SMG5	hHC026654	t2438042	chr1	-	156219019	156252630	0.71
SNX14	hHA035514	t2963313	chr6	-	86215217	86307586	-0.82

SPSB1	hHC022973	t2319252	chr1	+	9352942	9429580	0.77
SPTLC1	hHA035186	t3214582	chr9	-	94765899	94877690	-0.83
SRRT	hHC020421	t3015941	chr7	+	100472608	100486284	0.68
ST5	hHC009572	t3361971	chr11	-	8713154	8932493	0.68
TACC3	hHC020076	t2714955	chr4	+	1722818	1746895	0.60
TIMELESS	hHC011418	t3457824	chr12	-	56810167	56843192	0.63
TM7SF3	hHA034151	t3448481	chr12	-	27124527	27167327	-0.84
TMCC2	hHC013227	t2376376	chr1	+	205197324	205242465	0.76
TMEM132E	hHC019221	t3718236	chr17	+	32819166	32966327	0.72
TMEM49	hHA034602	t3764916	chr17	-	57915313	57917744	-0.86
TMTC3	hHC010424	t3425134	chr12	+	88536101	88593655	-0.82
TPCN2	hHC022004	t3337918	chr11	+	68816365	68929908	0.53
UBA2	hHR009662	t3829768	chr19	+	34919277	35003208	-0.87
UBE2N	hHC029873	t3465791	chr12	-	93802088	93861489	-0.65
UBTD1	hHC024134	t3259888	chr10	+	99258689	99330958	0.67
XRCC3	hHC022265	t3580832	chr14	-	104158599	104181796	0.64
ZBTB33	hHC009280	t3989089	chrX	+	119378531	119392247	-0.83
ZDHHC21	hHR005601	t3199431	chr9	-	14546162	14693787	-0.70
ZKSCAN5	hHC015797	t3014855	chr7	+	99086950	99132313	0.68
ZMAT3	hHC013459	t2706791	chr3	-	178734260	178789572	-0.88
ZNF473	hHR013607	t3839142	chr19	+	50528999	50556664	0.59
ZNF579	hHC014534	t3871557	chr19	-	56075948	56098806	0.79
ZNF672	hHC021765	t2390489	chr1	+	249132100	249143713	0.74

Supplementary Table 2. Weights assigned to each genotype of each SNP when computing the PCI.

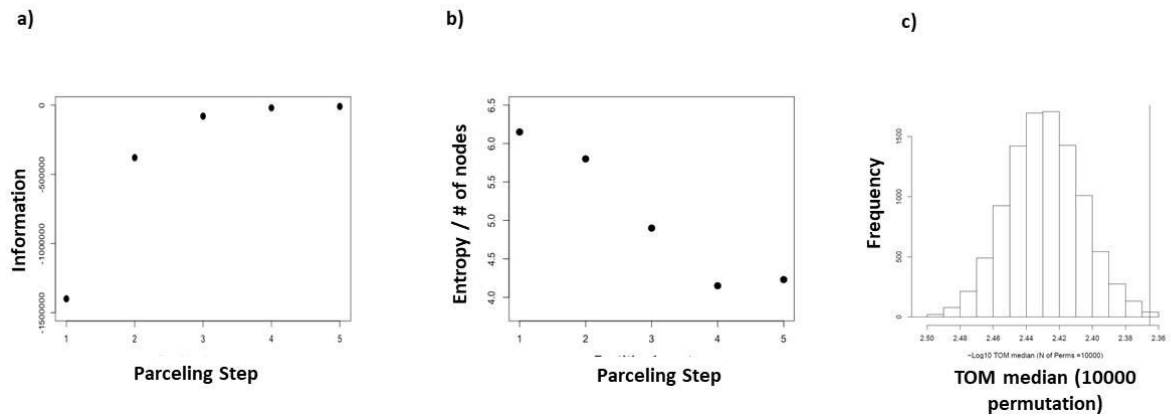
SNP	Genotype	Weight
rs7487813	GG	0.5
rs7487813	T-carriers	0.682083353
rs2267844	CC	0.763360813
rs2267844	CT	0.512798454
rs2267844	TT	0.5
rs663208	CC	0.765595704
rs663208	CT	0.613451133
rs663208	TT	0.5
rs17005918	C-carriers	0.353816908
rs17005918	TT	0.5
rs13101217	AA	0.466639623
rs13101217	AG	0.316876696
rs13101217	GG	0.5
rs1859464	CC	0.5
rs1859464	T-carriers	0.631121557
rs2278214	CC	0.5

rs2278214	CT	0.575577855
rs2278214	TT	0.782451245
rs7915524	CC	0.719045851
rs7915524	CT	0.487986131
rs7915524	TT	0.5
rs12509826	CC	0.234989278
rs12509826	CT	0.482600588
rs12509826	TT	0.5
rs10134399	CC	0.5
rs10134399	CT	0.622919329
rs10134399	TT	0.698026883
rs10906841	AA	0.706580205
rs10906841	AG	0.530555236
rs10906841	GG	0.5
rs2306251	AA	0.5
rs2306251	AG	0.633938465
rs2306251	GG	0.61574862
rs11602122	CC	0.5
rs11602122	CT	0.602444381
rs11602122	TT	0.331436767

Supplementary Table 3. Specification of the single nucleotide polymorphisms (SNPs) included in *DRD1*-PCI calculation. Abbreviations: MAF, minor allele frequency.

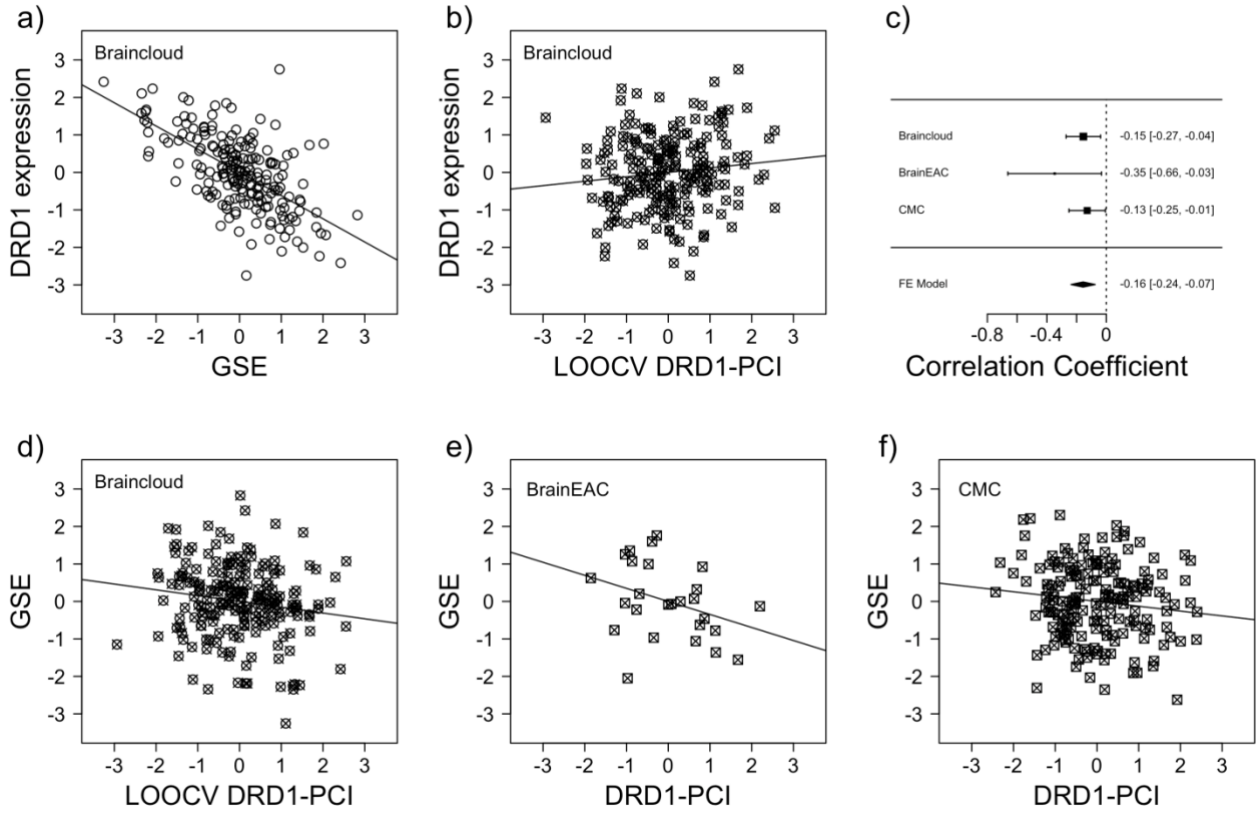
Rank	SNP	Locus	Module Gene	Gene name	Position	MAF	Type of variant
1	rs7487813	12q22	***UBE2N***	Ubiquitin Conjugating Enzyme E2 N	chr12:92311700	0.18	Intron variant
2	rs2267844	3p21.31	*SLC26A6*	Solute Carrier Family 26 Member 6	chr3:485665000	0.23	Intron variant (*PFKFB4*)
3	rs663208	11q14.2	*CCDC81*	Coiled-Coil Domain Containing 81	chr11:85870900	0.23	Intron variant (*ME3*)
4	rs17005918	4q31.1	***SCOC***	Short Coiled-Coil Protein	chr4:141584000	0.19	3 prime UTR variant (*RP11-542P2.1*)
5	rs13101217	3q21.1	***SEC22A***	SEC22 Homolog A, Vesicle Trafficking Protein	chr3:124449000	0.35	Intron variant
6	rs1859464	14q24.2	*PCNX1*	Pecanex Homolog (Drosophila)	chr14:70357600	0.14	Intergenic variant
7	rs2278214	2q11.2	***MGAT4A***	Mannosyl (Alpha-1,3-) - Glycoprotein Beta-1,4-N-Acetylglucosaminyl transferase, Isozyme A	chr2:985022000	0.21	Non coding transcript exon variant (*INPP4A*)
8	rs7915524	10p13	*FAM171A1*	Family With Sequence Similarity 171 Member A1	chr10:15333900	0.19	Intron variant
9	rs12509826	4q31.1	***SCOC***	Short Coiled-Coil Protein	chr4:141601000	0.25	Intron variant (*RP11-542P2.1*)
10	rs10134399	14q32.33	*XRCC3*	X-Ray Repair Cross Complementing 3	chr14:103153000	0.31	Intron variant (*RP11-73M18.2*)
11	rs10906841	10p13	*FAM171A1*	Family With Sequence Similarity 171 Member A1	chr10:15243300	0.34	Intron variant (*NMT2*)
12	rs2306251	4p16.3	*GAK*	Cyclin G Associated Kinase	chr4:776401	0.25	Splice region variant (*CPLX1*)
13	rs11602122	11q13.3	*TPCN2*	Two Pore Segment Channel 2	chr11:68537600	0.22	Intron variant (*RP11-554A11.6*)

Supplementary Figure 1



Identification of the DRD1 gene community. Panel a) represents the information content of the *DRD1* gene communities obtained with iterative partitions based on the average betweenness of the nodes in Braincloud. Note that information plateaus after three-four cycles. Panel b) shows entropy per node variation corresponding to each iteration in Braincloud. The minimum entropy is found after four cycles. Panel c) is a histogram of resampled topological overlap matrices based on random genes in BrainEAC and the vertical continuous line on the right shows the topological overlap of the genes in the *DRD1* community obtained following the procedure illustrated in panels a) and b). The graph is cut on the x-axis because the overlap is off-scale with respect to random resampling. Thus, the overlap is greater than chance in the replication microarray post mortem dataset.

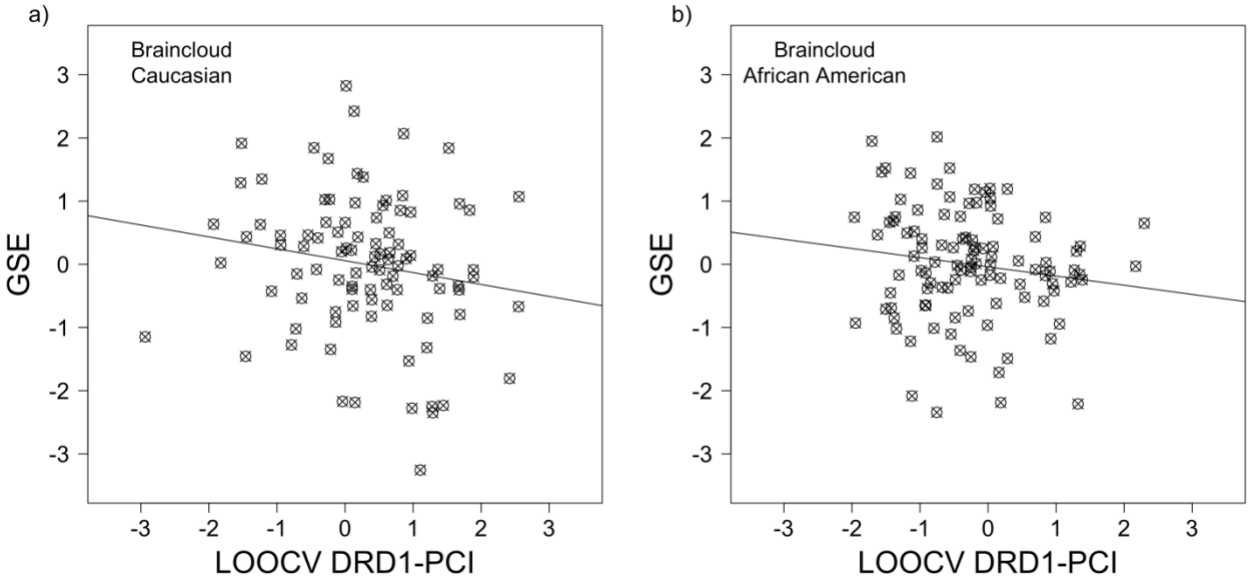
Supplementary Figure 2



Associations between the *DRD1*-PCI, the Gene Set Eigengene (GSE), and *DRD1* expression. Trendlines represent the linear regression fits in panels a-b) and d-e-f). All units in the scatter plots represent Z-scores of the variables indicated on the axes. a) This panel illustrates the inverted relationship between *DRD1* expression and the GSE In Braincloud. It is worth noting that the *DRD1*-PCI was defined in Braincloud as an index negatively correlated to the GSE. b) This panel illustrates the relationship between observed *DRD1* expression and the leave-one-out cross validated PCI (LOOCV *DRD1*-PCI) in Braincloud. c) Forest plot shows the size of the correlation between the *DRD1*-PCI and the GSE in subjects independent from the training set (the LOOCV *DRD1*-PCI was used for Braincloud). The diamond represents the fixed effect (FE) meta-analysis effect size. d) The X-axis depicts the LOOCV *DRD1*-PCI, whereas the Y-axis reports the observed GSE in Braincloud (the training dataset). e) Correlation observed in BrainEAC between the *DRD1*-

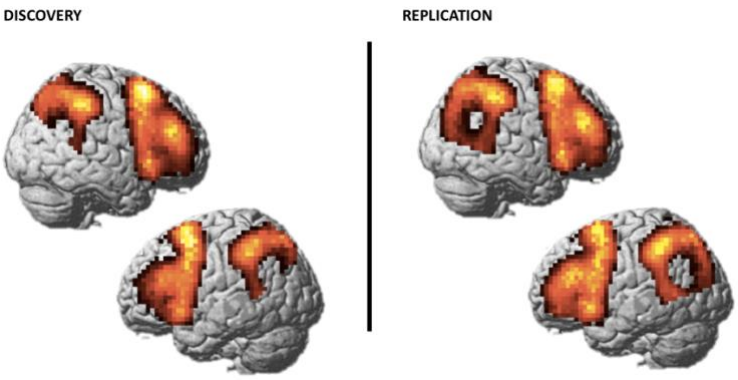
PCI and the GSE. f) Correlation observed in CommonMind Consortium (CMC) data between the *DRD1*-PCI and the GSE.

Supplementary Figure 3



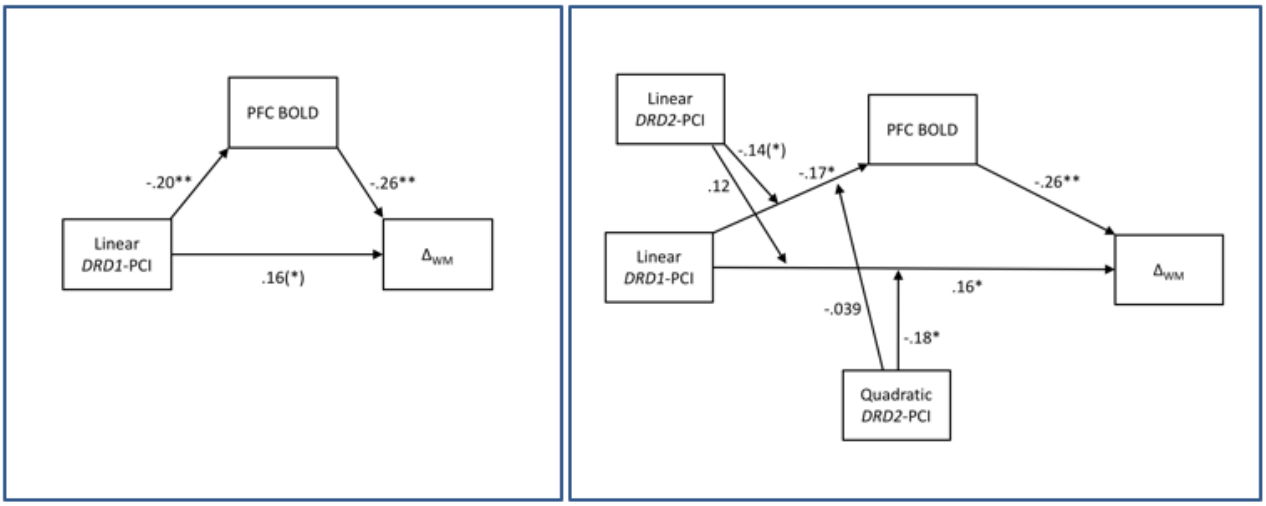
Correlation between the Gene Set Eigengene (GSE) and leave-one-out cross validated PCI (LOOCV *DRD1*-PCI) in Braincloud subjects of Caucasian and African American ancestry. The scatterplot illustrates the association of the LOOCV *DRD1*-PCI with the GSE in both ethnicities.

Supplementary Figure 4



Activation maps of WM-related activity. Left: main effect of task (conjunction null(1,2,3 back) > baseline) for the discovery sample ($p<0.05$). Right: main effect of task (WM (2 back) > baseline) for the replication sample ($p<0.05$)

Supplementary Figure 5



Mediation analysis models and results. Left: discovery sample, mediation model. Right: discovery sample, mediation/moderation model. Statistical significance is represented as follows: (*) = p < .1; * = p < .05; ** = p < .01.